Mol Genet Genomics (2007) 277:357–364
DOI 10.1007/s00438-006-0194-9

ORIGINAL PAPER

Mutator-like elements identified in melon, Arabidopsis and rice contain ULP1 protease domains

Hans van Leeuwen · Amparo Monfort · Pere Puigdomenech

Received: 21 July 2006 / Accepted: 5 November 2006 / Published online: 29 November 2006

Abstract The transposon *Mutator* was first identified in maize, and is one of the most active mobile elements in plants. The *Arabidopsis thaliana* genome contains at least 200 *Mutator*-like elements (MULEs), which contain the *Mutator*-like transposase gene, and often additional genes. We have detected a novel type of MULEs in melon (CUMULE), which, besides the transposase, contains two ubiquitin-like specific protease-like sequences (ULP1). This element is not present in the observed location in some melon cultivars. Multiple copies of this element exist in the *Cucumis melo* genome, and it has been detected in other Cucurbitaceae species. Analysis of the *A. thaliana* genome revealed more than 90 CUMULE-like elements, containing one or two *Ulp1*-like sequences, although no evidence of mobility exists for these elements. We detected various putative transposable elements containing ULP1-like sequences in rice. The discovery of these MULEs in melon and Arabidopsis, and the existence of similar elements in rice and maize, suggest that a proteolytic function may be important for this subset of the MULE transposable elements.

Keywords Transposons · Melon · *Mutator* · Comparative analysis · Protease

Communicated by M.-A. Grandbastien.

Electronic supplementary material Supplementary material is available in the online version of this article at http://dx.doi.org/10.1007/s00438-006-0194-9 and is accessible for authorized users.

Nucleotide sequence data reported are available in the GenBank database under the accession number AY524004.

H. van Leeuwen · P. Puigdomenech (✉)
Departament de Genètica Molecular, Laboratori de Genètica Molecular Vegetal, CSIC-IRTA, Jordi Girona 18, 08034 Barcelona, Spain
e-mail: pprgmp@ibmb.csic.es

A. Monfort
Departament de Genètica Vegetal. Ctra. de Cabrils s/n, Laboratori de Genètica Molecular Vegetal, CSIC-IRTA, 08348 Cabrils, Barcelona, Spain

Introduction

The transposon *Mutator* was first identified in maize by Robertson (1978) as a highly efficient system of mutagenesis. It is recognized as being one of the most active mobile elements in plants and has been widely employed as a mutagenizing system. It is also the basis for the first reverse genetic systems used in cereals (Meely and Briggs 1995). Many *Mutator* elements have been identified and sequenced (Bennetzen 1996). They are very heterogeneous and they have been classified into different subgroups according to their length and transposition properties. However, important features of the biology of these elements remain elusive.

Comparison of the *Mutator* elements has shown that they contain terminal inverted repeats (TIRs), approximately 200 bp long. The autonomous *MuDR* element in maize encodes two genes. One of these genes (*mudrA*) has been shown to code for a transposase (MURA), while the other gene (*mudrB*) encodes a protein (MURB) that still has no function attached to its sequence. The *mudrA* gene appears to be essential for the transposition activity of the element, while deletions in *mudrB* may interfere with the re-integration process (Lisch et al. 1999). While MURA has a nuclear localization it is not completely clear for

MURB and the two proteins do not seem to interact in two hybrid assays (Ono et al. 2002).

The *Arabidopsis thaliana* (Arabidopsis) genome contains at least 200 individual ORFs related to the MURA transposase of the maize *MuDR* element (Le et al. 2000; Singer et al. 2001). Of these elements, also called MULEs for *Mutator*-like element (Le et al. 2000), only 22 were flanked by TIRs (Singer et al. 2001), having a structure similar to the *MuDR* element. However, none of these elements contain a sequence that resembles the *mudrB* gene. In the rice genome more than 8,000 MULEs have been detected of which 24% contain sequences with similarity to coding regions of predicted genes that are unrelated to transposons (Juretic et al. 2005). This phenomenon of transduplication of gene fragments had previously been described in maize (Talbert and Chandler 1988), Arabidopsis (Le et al. 2000; Yu et al. 2000) and rice (Turcotte et al. 2001).

Cucumis melo has a relatively compact genome (Arumuganathan and Earle 1991), and it has already been shown that a degree of synteny of melon with Arabidopsis is to be expected (van Leeuwen et al. 2003). In the course of studying the genomic structure in a region of *Cucumis melo*, rich in homologues to resistance genes (van Leeuwen et al. 2005), a sequence with the features of a MULE transposon was found and compared to the Arabidopsis and maize elements. This element, named CUMULE, contains sequences related to the ULP1 protease besides the MURA-like transposase. The presence of the CUMULE-like elements in various species indicates a role for the ULP1 proteases in the life cycle of this subset of MULEs.

Materials and methods

Sequence analysis

The 31O16 BAC sequence (van Leeuwen et al. 2005) was analyzed for localization of repeated sequences with the JDotter software (Brodie et al. 2004). The Arabidopsis genome sequence was screened with the TBLASTN program (Altschul et al. 1997) at the TAIR BLAST web site (http://www.arabidopsis.org/Blast/) with the melon MURA (GenBank AAU04773), ULP1-1 (GenBank AAU04774) and ULP1-2 (GenBank AAU04775) sequences, and the Arabidopsis protein sequences AT2G05690 (MURA-like), AT5G45570 (ULP1-like) and ESD4 (ULP1-like) with the parameters: AGI whole genome (BAC clones) (DNA), no filter and an arbitrarily chosen, E-value of 0.001. CUMULE-like sequences in the *Oryza sativa* genome were searched for with the WU-BLAST 2.0 program (TBLASTN) at http://www.tigrblast.tigr.org/eukblast/index.cgi?project=osa1 using the melon CUMULE ULP1 sequences and the proteins encoded by the *MuDR* and *Jittery* transposons from *Zea mays* against all rice BAC and PAC sequences in GenBank.

Phylogenetic analysis was performed with the MURA-like and ULP1-like sequences from the putative elements detected in Arabidopsis. Pseudogenes and nonannotated sequences were excluded from the analysis (see S1). Multiple protein sequence alignment was performed with the ClustalW program version 1.82 (Thompson et al. 1994) at the EBI web site, with parameters gapopen 10, gapext 0.05 and score matrix Gonnet 250. For the MURA alignment the region from position 235 to 396 (numbering in the *Cucumis melo* MURA protein) was selected, which is inside the conserved MURA transposase domain as detected by RPS-BLAST at NCBI (Marchler-Bauer and Bryant 2004). For the ULP1 alignment the region from position 101 to 211 (numbering in the *Cucumis melo* MURA protein) was selected, which is outside the conserved ULP1 domain as detected by RPS-BLAST at NCBI (Marchler-Bauer and Bryant 2004). Bootstrapping (1,000×), calculation of the distance matrix and tree building were performed with the seqboot, protdist (JTT matrix), neighbor (NJ method) and consense programs of the Phylip 3.6 package (Felsenstein 1989, 2004), respectively.

PCR experiments on CUMULE elements

PCR primers were designed external and internal of the 5′ and 3′ ends of the CUMULE sequence: E5: 5′-TTT ACC TCT CAC TTG CAC ATC-3′, E3: 5′-TGC TAC CCA CTT CAA CAT ATC-3′ and I3: 5′-GAT GTC ATG ATT CAA GCC AAG-3′. Genomic DNA of several Cucurbitaceae species (*Cucumis africanus*, *Cucumis prophetarum*, *Cucumis globosus*, *Cucumis metuliferus*, *Cucumis sativus*, *Citrullus lanatus* and *Cucurbita pepo*) and melon (*Cucumis melo* L.) varieties (‘Piel de Sapo’, ‘Songwhan Charmi’, ‘Snake cucumber’, ‘Freeman Cucumber’, ‘G 22841’ and ‘Ein Dor’) (Monforte et al. 2003) was used as template. PCR experiments were performed in 15 µl with 2.1 mM MgCl2, 0.1 mM dNTPs, 0.25 mM primers and 1 Unit Taq DNA polymerase (Roche Diagnostics S.L., Spain). PCR cycling conditions with the E3–I3 primer combination were as follows: 94°C for 1 min followed by 35 cycles at 94°C for 30 s, 53°C for 30 s and 72°C for 1 min) and a final cycle at 72°C for 7 min. With the E5–E3 primer combination the conditions were: 94°C for 1 min followed by 35 cycles at 94°C for 30 s, 50°C for

30 s and 72°C for 1 min) and a final cycle at 72°C for 7 min. Amplified sequences of the expected size were purified with the GFX™ PCR DNA and Gel Band Purification Kit (Amersham Pharmacia Biotech Inc., USA) and sequenced with the Terminator cycle sequencing ready reaction kit (Perkin Elmer), using the corresponding primers. The sequencing reactions were visualized and processed with ABI PRISM377.

PCR primers at the CUMULE 5′ and 3′ end of the 5′ end TIR were used for amplification using the DNA pools from the PIT92 BAC library (Yu et al. 2000) as template: 5′TIR-F: 5′-TTG GGA AAT TGC CAA AAA TAG G-3′ and 5′TIRR: 5′-CAG CAG ATA AAT TGA AAA TCG G-3′. PCR experiments were performed in 25 µl with 1.5 mM MgCl2, 0.1 mM dNTPs, 0.25 mM primers and 1 Unit Taq DNA polymerase (Roche Diagnostics S.L., Spain). PCR cycling conditions were as follows: 94°C for 2 min followed by 35 cycles at 94°C for 20 s, 62°C for 20 s and 72°C for 25 s) and a final cycle at 72°C for 5 min.

Results

Discovery of CUMULE

Analysis of the melon 31O16 BAC sequence (van Leeuwen et al. 2005) produced, among other gene sequences, a putative protein with high similarity to the *MuDR* family of transposases (conserved protein family domain pfam03108), as detected with RPSBLAST at NCBI (Marchler-Bauer and Bryant 2004) (E-value of 3E-33 with an alignment coverage of 87.7%). Full length elements (*MuDR*) in maize encode two proteins: the MURA protein with a transposase function, and the MURB protein of unknown function. The two genes encoding these proteins are in opposite orientation with the 3′ ends in the center of the transposable element. The coding region of the element is usually flanked by TIRs. An initial analysis of the melon sequence downstream of the *mudrA*-like gene did not show a *mudrB*-like gene, nor were TIRs detected. The genes situated downstream of the *mudrA*-like melon sequence encode a protein (with an uninterrupted open reading frame) with significant similarity to the ULP (Ubiquitin-like specific protease) family of proteases (E-value 6E-12). The two melon *Ulp1*-like genes (*Ulp1-1* and *Ulp1-2*) encode the conserved protein family domain pfam02902 (Peptidase_C48, Ulp1 protease family) as detected with RPS-BLAST at NCBI (Marchler-Bauer and Bryant 2004) (E-values of 2E-17 and 2E-09, respectively, and alignment coverages of 70.1 and 78.2%, respectively). Possible TIRs with lengths of 425 and 395 bp, respectively, with 67% identity between them were found with one copy upstream of the *mudrA*-like gene and the other copy downstream of the second *Ulp1*-like gene. The TIRs consist of two highly conserved regions separated by a less conserved region. A 5-bp perfect target site duplication flanks these TIRs, which brings the total length of this element to 12,089 bp (Fig. 1). The DNA sequence of this element (CUMULE; Cucumis *Mutator*-like transposable element) has been deposited in the GenBank database [GenBank: AY524004]. Sequence analysis of the melon *Ulp1* genes shows homology with *Ulp1*-like genes from a wide variety of species, including vertebrates (Fig. 2). We did not detect any similar *Cucumis melo* sequences in the sequence databases.

Copy number and mobility of CUMULE

We designed primers based on the 5′ TIR to amplify possible additional copies of CUMULE in the *Cucumis melo* PIT92 genome. PCR amplification with ten DNA pools (each containing 384 BAC clones) of the PIT92 BAC library (van Leeuwen et al. 2003) gave a fragment of the expected size in all pools, and some pools contained additional fragments of a slightly different size. Using the pool columns from two pools (each containing 32 BAC clones) also amplified a fragment of the expected size in all samples. Fragments were purified and sequenced, and alignment of the sequences showed high identity with various minor changes or short deletions and insertions (results not shown). The method used in this analysis did not allow distinguishing multiple copies per pool. Based on the characteristics of the PIT92 BAC library (van Leeuwen et al. 2003), with a 6.2 times coverage of the genome, these data suggest that at least, but most likely more, 120 CUMULE elements are present in the *Cucumis melo* genome.

To detect the CUMULE element in the genome of other Cucurbitaceae species and other melon cultivars,

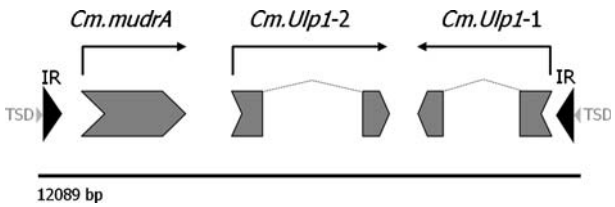

Fig. 1 The melon CUMULE element. Schematic representation of the element with MURA-like transposase, ULP1-like proteases, inverted repeats (*IR*) and target site duplications (*TSD*). The 'gaps' in the *Ulp1*-like genes stand for the introns. The full sequence plus annotation has been deposited in GenBank (AY524004)

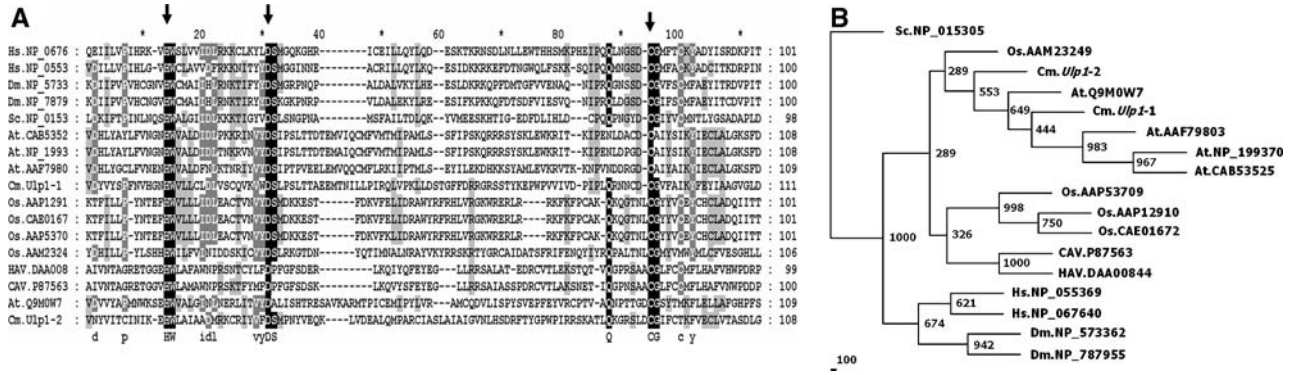

Fig. 2 Protein sequence analysis of a highly conserved region of the ULP1 protease. **a** ClustalW protein alignment. *Arrowheads* mark likely residues of the catalytic triad in the ULP1 cysteine protease family. **b** Phylogenetic tree. Proteins were aligned with ClustalW, and the PHYLIP software (seqboot, protdist, neighbor and consense) was used to calculate a consensus tree. Bootstrapping was performed with 1,000 replicates. The tree was visualized with TREEVIEW software. Accession numbers follow the species abbreviations, At, *Arabidopsis thaliana*; *CAV*, Canine adeno virus; Cm, *Cucumis melo*; Dm, *Drosophila melanogaster*; *HAV*, Human adeno virus; Hs, *Homo sapiens*; Os, *Oryza sativa*; Sc, *Saccharomyces cerevisiae*

we designed PCR primers at the 5′ end and the 3′ end of the CUMULE element in the melon BAC 31O16. Primers were designed inside (I3) and outside (E5 and E3) the transposable element. In *Cucumis globosus, Cucumis melo* L. 'Ein Dor', *Cucumis melo* L. 'Songwhan Charmi' and *Cucumis melo* L. 'Piel de Sapo' (this last variety is a parental of the PIT92 line in which CUMULE was discovered) we amplified a fragment of the expected size with the E3–I3 primer combination. Sequencing confirmed the presence of a CUMULE-like element in *Cucumis melo* L. 'Songwhan Charmi' and *Cucumis melo* L. 'Piel de Sapo'. To verify that the CUMULE elements are active, two primers (E5–E3) from the external sequence of CUMULE were used. The E5–E3 PCR should only amplify a sequence of approximately 375 bp in species where the CUMULE element is absent at this particular site of the genome. The 375-bp band was amplified in the melon cultivar *Cucumis melo* L. 'G 22841', from Senegal, and sequencing of this band confirmed the absence of the CUMULE element at this particular region of its genome. The PCR with the E3 and I3 primers was negative for this cultivar. We did not detect a target site duplication in the sequenced fragment, which is usually left when transposons excise themselves from the genome.

The CUMULE transposon family in Arabidopsis

We performed a detailed *in-silico* analysis on the complete Arabidopsis genome searching for pairs of *mudrA*-like and *Ulp1*-like genes. We limited our search to pairs that spanned a region not more than 20 kb, with the two genes in opposite direction. The TBLASTN program was used at the TAIR website, with the MURA (GenBank AAU04773), ULP1-1 (GenBank AAU04774) and ULP1-2 (GenBank AAU04775) sequences from melon, and the Arabidopsis protein sequences AT2G05690 (MURAlike), AT5G45570 (ULP1-like) and ESD4 (ULP1-like). We detected 164 *Ulp1*-like sequences with an arbitrarily chosen, E-value threshold of 0.001. The *ESD4*-like genes in the Arabidopsis genome were significantly different at the sequence level compared to the other *Ulp1*-like genes in the genome, and they were not part of putative CUMULE elements.

With the TBLASTN matches, we then manually checked the Arabidopsis annotation at the TAIR web site to compose a detailed list of all the detected CUMULE-like elements in Arabidopsis (S1). There were 91 putative elements distributed over all five chromosomes containing both *mudrA*-like and *Ulp1*-like genes, resembling the element from melon (S1). Most of these elements consisted of one *mudrA*-like gene and one *Ulp1*-like gene (AtCUMULE-I, 80 elements), but some had two *Ulp1*-like genes, as in the melon BAC (AtCUMULE-II, 11 elements) although they were in the same direction, thus different from the orientation in the melon element. Thus, from the 164 *Ulp1*-like sequences that we detected in the Arabidopsis genome, 102 (62%) were associated with *mudrA*-like sequences. Of these 102 sequences, 13 encoded pseudogenes. The average size of the elements in Arabidopsis is 11.3 kb for the AtCUMULE-I elements and 14.3 kb for the AtCUMULE-II elements, which is similar to the size of CUMULE in melon, and represent approximately 1,061 kb (0.7%) of the genome. The elements are distributed throughout the genome with a higher number located near the centromeres (Fig. 3). A BLASTN analysis of the Arabidopsis genome with the melon

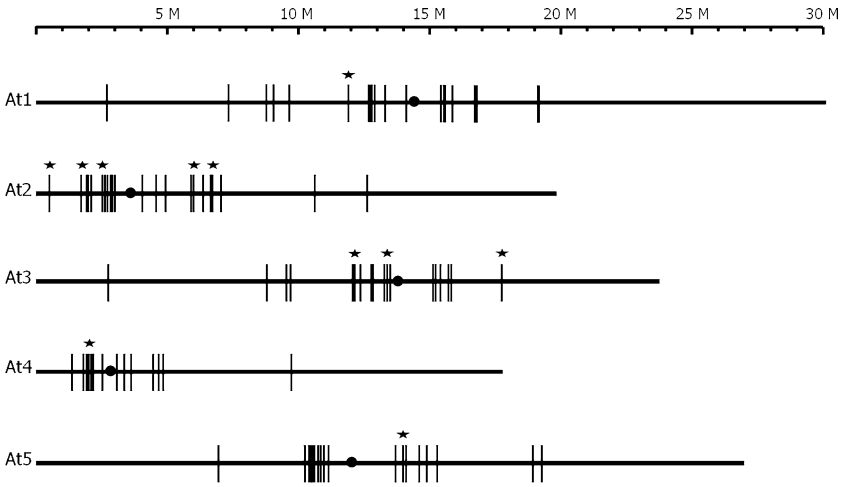

Fig. 3 Distribution of the AtCUMULE-I and AtCUMULE-II elements in the *Arabidopsis thaliana* genome. AtCUMULE-II elements are indicated with an *asterisk*. The centromeres of the chromosomes are represented as *black dots*

CUMULE TIRs did not retrieve any similar sequences. We analyzed the flanking regions of the At-CUMULE elements for TIRs but could only detect short putative TIRs for some elements (data not shown).

Phylogenetic analysis of the MURA-like sequences and ULP1-like sequences from the AtCUMULE-I and AtCUMULE-II elements (Fig. 4) resulted in several subfamilies. The members of these subfamilies grouped together in corresponding groups in the MURA sequence analysis and the ULP1 sequence analysis. According to these results the AtCUMULE-I elements can be divided into six subfamilies, while the AtCUMULE-II elements all grouped in one subfamily. The ULP1 analysis shows that the ULP1-1 sequences, located at the extremes of the elements, all grouped together, as well as the ULP1-2 sequences, located in the center of the elements. This suggests an early duplication event of the *Ulp1*-like gene in one At-CUMULE-I element, after which the element duplicated and spread through the genome.

We detected three annotated genes in the Arabidopsis genome encoding proteins with the transposase, and ULP1 domain (At3g42470, At4g05550 and At5g37330), a situation that we also detected in the rice genome (described below).

Additional genes, besides the *mudrA* and *Ulp1*-like genes, were found inside a number of CUMULE elements in Arabidopsis, e.g., in 15 At-CUMULE-1 elements we detected genes encoding a protein containing the InterPro (Mulder et al. 2003) IPR003871 domain that has an unknown function (see S1). It is interesting to note that the phylogenetic analysis shows that five of the six elements from the yellow clade contain genes with the IPR003871 domain (Fig. 4).

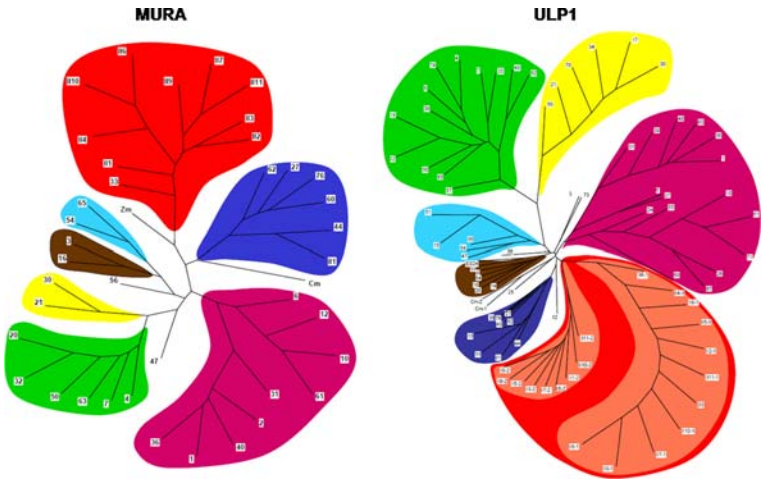

Fig. 4 Phylogenetic trees for the ULP1-like and MURA-like sequences from the AtCUMULE-I and AtCUMULE-II elements from *Arabidopsis thaliana*. In the analysis, the MURA protein from *Zea mays* and from *Cucumis melo* was included. Phylogenetic analysis and sequences (only complete sequences without stop codons were considered) are described in Materials and methods. The putative elements of which both the *MuDRA* and *Ulp1* genes were pseudogenes or not annotated are not included in this analysis (see S1). Corresponding subfamilies in the two trees are visualized with manually assigned colors

CUMULE-like sequences in *Oryza sativa* and *Zea mays*

Analysis of the rice genome with the TBLASTN program using the ULP1-like proteins from CUMULE resulted in various putative transposable elements containing the ULP1-like protease. Initial analysis gave four sequences encoding proteins containing a transposase domain (tnp2-like) and a ULP1 domain (Fig. 5, *Oryza sativa*, C). We then analyzed the maize genome which contains multiple *Mutator* elements with the *mudrA* transposase gene and the *mudrB* gene. No *Ulp1*-like genes were found in the maize sequences currently available in the databases. An active MULE, *Jittery*, has been found recently (Xu et al. 2004). It does not contain the *mudrB* gene, and its MURA-like transposase contains an additional domain (FAR1; pfam PF03101). A second analysis of the rice genome using the *Jittery* sequence from maize detected multiple sequences similar to *Jittery*, and many contained an additional domain similar to the ULP1 protease domain (Fig. 5, *Oryza sativa*, D). Protein sequence alignment of the rice sequences and the *Jittery* sequence shows that there is low homology in the region of the ULP1 domain in the rice sequences. Using the ULP1 domain from the rice *Jittery*-like sequences, a third analysis of the rice genome gave nine CUMULE-like elements. These elements contain a sequence encoding the FAR1 and MURA domains and one or two sequences encoding a ULP1-like protease (Fig. 5, *Oryza sativa*, A and B).

Discussion

We have discovered a new transposon family, CUMULE, in a resistance gene cluster of the melon genome. This family is structurally similar to the *Mutator* transposon family of maize, and the CUMULE transposase is significantly similar to the maize *Mutator* transposase. The other two genes in CUMULE encode ULP1 proteases that are homologous to adenoviral and yeast proteases (Fig. 2). The ULP1 protease belongs to a class of the family of cysteine proteases (Barrett and Rawlings 2001) represented by adenain, an endopeptidase from the double stranded DNA viruses of the adenovirus group. This is illustrated in Fig. 2 in the alignment of the sequences with the protease sequences deduced from two adenoviruses from mammal species. In adenovirus, the protease is required for processing of viral precursor proteins during virion maturation, but is also involved in de-ubiquitination activity of cellular and/or

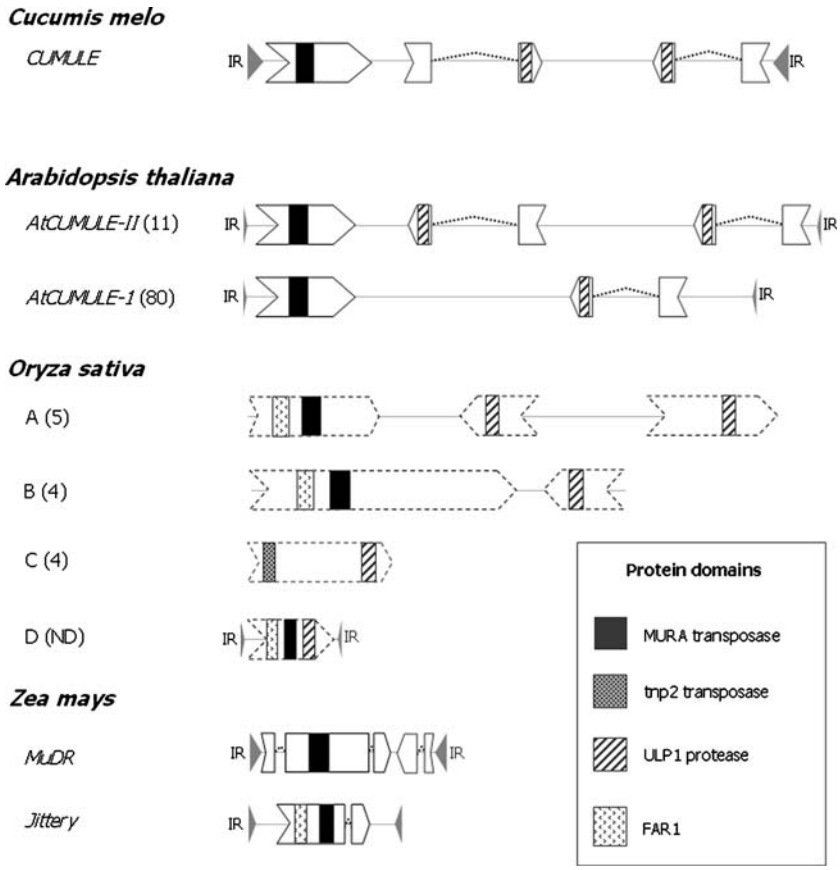

Fig. 5 Structure of CUMULE-like elements in different species. *Cucumis melo*: CUMULE (acc. no. AY524004); *Arabidopsis thaliana*: 11 AtCUMULEII elements (represented here AtCUMULE-II11), 80 AtCUMULE-I elements (represented here AtCUMULE-I80); *Oryza sativa*: discontinued lines in gene representation means that no information on introns is available. *A* Five elements, BACs OSJNBa0074L08, OSJNBa0088A01, OSJNBb0013K08, OSJNBb0062H02, P0705B06; *B* four elements, BACs B1151D08, B1168H06, Bb0115K12, OSJNBa0050F15; *C* four elements, genes AAM23249, AAP53709, CAE01672, AAP12910; *D* number of elements not determined, e.g. gene OSJNBa0035H01.9; *Zea mays*: *MuDR* (acc. no. M76978), *Jittery* (acc. no. AF247646)

viral proteins (Balakirev et al. 2002). This last function could, hypothetically, play a role in the integration process of the excised transposon sequence into a new location in the genome, in a similar way, as described for bacteriophage Mu. In *Escherichia coli*, a member of the Clp/HSP100 ATPase unfoldase/protease family is involved in remodeling of the transpososome (recombination complex) from bacteriophage Mu, and is involved in the integration of the bacteriophage Mu DNA into the host genome (Jones et al. 1998). And it has been shown that ULP proteases take part in the regulatory pathway involving SUMOylation (Kurepa et al, 2003) in Arabidopsis. Recently, the involvement of ULP proteases in transposon action has been proposed (Hoen et al. 2006).

We detected additional copies of CUMULE in the PIT92 melon accession by amplifying the TIR in pools of the BAC library. From these results, we can conclude that there are at least 120 copies in melon, a number comparable to the 91 copies we detected in the Arabidopsis genome. Further analysis should give a more precise estimation of the copy number of this new transposon family in the *Cucumis melo* genome. The CUMULE transposon was detected, using PCR, in *Cucumis globosus* and *Cucumis melo* L. 'Songwhan Charmi'. The absence of the CUMULE element in *Cucumis melo* L. 'G 22841' at the site homologous to the location in our BAC 31O16, and the absence of a duplicated target site suggests that CUMULE has never been present in this location in this accession, and that CUMULE is an active transposon in the melon genome. Alternatively, the element has not left a target site duplication upon excision.

Analysis of the Arabidopsis genome revealed the presence of at least 91 CUMULE-like elements. These elements group into two classes: AtCUMULE-I with one *Ulp*1-like gene, and AtCUMULE-II with two *Ulp*1-like genes (Fig. 5). Of the 164 *Ulp*1-like sequences that we detected in the Arabidopsis genome, 62% were associated with *Mutator*-like transposase sequences, using our criteria. The Arabidopsis *ESD4* gene is the only *Ulp*1-like gene in plants to which a function has been assigned. It encodes a ULP1 protease involved in flowering-time regulation in Arabidopsis (Murtas et al. 2003). Eighty-seven percent of the *Ulp*1-like sequences in the AtCUMULEs in Arabidopsis encoded ORFs. Phylogenetic analysis of both the *Mutator*-like transposase and the ULP1 protease suggest that the genes encoding these sequences have evolved together in the Arabidopsis genome. The analysis also suggests that the AtCUMULE-II elements evolved from a *Ulp*1-like gene duplication in a single AtCUMULE-I element.

The presence of additional genes in a subset of the AtCUMULEs indicate that these elements may take an active role in transduplication of gene fragments across different regions of the plant genomes as it has already been described in maize (Talbert and Chandler 1988), Arabidopsis (Le et al. 2000; Yu et al. 2000) and rice (Turcotte et al. 2001). In rice, this mechanism has been suggested to be involved in the creation of novel protein-coding genes (Jiang et al. 2004), although a recent study in rice has provided evidence that the gene duplication events caused by MULEs create mainly pseudogenes (Juretic et al. 2005). The Pack-MULES described by Jiang et al. (2004) all have TIRs belonging to the same MULE family. Although we detected putative TIRs in the melon CUMULE element, the inability to detect TIRs in the majority of the Arabidopsis CUMULE-like elements made it impossible to determine whether there are non-autonomous elements of this family in the Arabidopsis genome. Taking these findings into account, the CUMULEs are most likely not a subset of the Pack-MULES family.

In rice, we detected several sequences that code for proteins containing a FAR1 domain, a MURA-like transposase domain and a ULP1-like domain (Fig. 5, *Oryza sativa*, D). These sequences are similar to the *Jittery* transposon (Xu et al. 2004) of *Zea mays*, except for the absence of the ULP1 domain in the latter. In rice, we also detected genes encoding proteins with the FAR1 and MURA-like transposase domains next to genes encoding ULP1 proteases, resembling the structure of CUMULE (Fig. 5, *Oryza sativa*, A and B). Finally, four putative transposons exist in rice, with a protein containing a MURA-like transposase domain and a ULP1-like domain (Fig. 5, *Oryza sativa*, C). The different types of elements detected in rice suggest that CUMULE-like elements exist in this species.

In conclusion we have detected a novel family of *MULE*s in melon, and subsequently detected multiple copies of similar elements in Arabidopsis and rice. Besides the MURA-like transposase sequence, this novel family of MULEs contains sequences related to the ULP1 protease family. The phylogenetic analysis of these elements in Arabidopsis supports the discovery of this novel type of MULEs, and shows two subgroups with either one or two copies of the ULP1-like sequence. The discovery of CUMULE in melon, and CUMULE-like elements in Arabidopsis, the existence of similar elements in rice, and the function of the ULP1-like proteins in other species, suggest that a proteolytic function may be important for this subset of MULE transposable elements.

Acknowledgments The authors thank Mrs. Mercè Miquel for DNA sequencing and Dr. J.M. Casacuberta for helpful discussions. This project was funded by grant 2FD97-0286-(02-01) from CICYT and grant 2000FI-00316 to Hans van Leeuwen from CIRIT. This work was carried out within the framework of the Centre de Referència de Biotecnologia, Generalitat de Catalunya.

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208–218

Balakirev MY, Jaquinod M, Haas AL, Chroboczek J (2002) Deubiquitinating function of adenovirus proteinase. J Virol 76:6323–6331

Barrett AJ, Rawlings ND (2001) Evolutionary lines of cysteine peptidases. Biol Chem 382:727–733

Bennetzen JL (1996) The *Mutator* transposable element system of maize. Curr Top Microbiol Immunol 204:195–229

Brodie R, Roper RL, Upton C (2004) JDotter: a Java interface to multiple dotplots generated by dotter. Bioinformatics 20:279–281

Felsenstein J (1989) PHYLIP—Phylogeny Inference Package (Version 3.2). Cladistics 5:164–166

Felsenstein J (2004) PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle

Hoen DR, Park KC, Elrouby N, Yu Z, Mohabir N, Cowan RK, Bureau TE (2006) Transposon-mediated expansion and diversification of a family of ULP-like genes. Mol Biol Evol 23:1254–1268

Jiang N, Bao Z, Zhang X, Eddy SR, Wessler SR (2004) Pack-MULE transposable elements mediate gene evolution in plants. Nature 431:569–573

Juretic N, Hoen DR, Huynh ML, Harrison PM, Bureau TE (2005) The evolutionary fate of MULE-mediated duplications of host gene fragments in rice. Genome Res 15:1292–1297

Jones JM, Welty DJ, Nakai H (1998) Versatile action of *Escherichia coli* ClpXP as protease or molecular chaperone for bacteriophage Mu transposition. J Biol Chem 273:459–465

Kurepa J, Walker JM, Smalle J, Gosink MM, Davis SJ, Durham TL, Sung DY, Vierstra RD (2003) The small ubiquitin-like modifier (SUMO) protein modification system in Arabidopsis. J Biol Chem 278:6862–6872

Le QH, Wright S, Yu Z, Bureau T (2000) Transposon diversity in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 97:7376–7381

Lisch D, Girard L, Donlin M, Freeling M (1999) Functional analysis of deletion derivatives of the maize transposon *MuDR* delineates roles for the MURA and MURB proteins. Genetics 151:331–341

Marchler-Bauer A, Bryant SH (2004) CD-Search: protein domain annotations on the fly. Nucleic Acids Res 32:W327–W331

Meeley RB, Briggs SP (1995) Reverse genetics for maize. Maize Genet Coop Newsl 69:67–82

Monforte AJ, Garcia-Mas J, Arús P (2003) Genetic variability in melon based on microsatellite variation. Plant Breeding 122:1–6

Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Barrell D, Bateman A, Binns D, Biswas M, Bradley P, Bork P, et al. (2003) The InterPro Database, 2003 brings increased coverage and new features. Nucleic Acids Res 31:315–318

Murtas G, Reeves PH, Fu YF, Bancroft I, Dean C, Coupland G (2003) A nuclear protease required for flowering-time regulation in Arabidopsis reduces the abundance of small ubiquitin-related modifier conjugates. The Plant Cell 15:2308–2319

Ono A, Kim SH, Walbot V (2002) Subcellular localization of MURA and MURB proteins encoded by the maize *MuDR* transposon. Plant Mol Biol 50:599–611

Robertson DS (1978) Characterization of a *Mutator* system in maize. Mutat Res 51:21–28

Singer T, Yordan C, Martienssen RA (2001) Robertson's *Mutator* transposons in *A. thaliana* are regulated by the chromatin-remodeling gene Decrease in DNA Methylation (DDM1). Genes Dev 15:591–602

Talbert LE, Chandler VL (1988) Characterization of a highly conserved sequence related to *Mutator* transposable elements in maize. Mol Biol Evol 5:519–529

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Turcotte K, Srinivasan S, Bureau T (2001) Survey of transposable elements from rice genomic sequences. Plant J 25:169–179

van Leeuwen H, Monfort A, Zhang HB, Puigdomènech P (2003) Identification and characterisation of a melon genomic region containing a resistance gene cluster from a constructed BAC library. Microcolinearity between *Cucumis melo* and *Arabidopsis thaliana*. Plant Mol Biol 51:703–718

van Leeuwen H, Garcia-Mas J, Coca M, Puigdoménech P, Monfort A (2005) Analysis of the melon genome in regions encompassing TIR-NBS-LRR resistance genes. Mol Genet Genomics 273:240–251

Xu Z, Yan X, Maurais S, Fu H, O'Brien DG, Mottinger J, Dooner HK (2004) Jittery, a *Mutator* distant relative with a paradoxical mobile behavior: excision without reinsertion. Plant Cell 16:1105–1114

Yu Z, Wright SI, Bureau TE (2000) *Mutator*-like elements in *Arabidopsis thaliana*. Structure, diversity and evolution. Genetics 156:2019–2031